

# Table des matières

<b>Préface</b> . . . . .	11
Jean-Charles POMEROL	
<b>Introduction</b> . . . . .	13
Marine CORLOSQUET-HABART et Jacques JANSSEN	
<b>Chapitre 1. Introduction au big data et application à l'assurance</b> . . . . .	17
Romain BILLOT, Cécile BOTHOREL et Philippe LENCA	
1.1. Le déluge des données : une journée-type des années 2010 . . . . .	17
1.2. Comment définir les big data ? . . . . .	20
1.3. Une caractérisation des mégadonnées à travers les 5 V . . . . .	21
1.3.1. Variété . . . . .	22
1.3.2. Volume . . . . .	23
1.3.3. Vitesse . . . . .	25
1.3.4. Vers les 5 V : véracité et valeur. . . . .	25
1.3.5. Autres V possibles . . . . .	26
1.4. Architecture. . . . .	26
1.4.1. Un écosystème technique de plus en plus complexe . . . . .	27
1.4.2. Migration vers une stratégie orientée donnée. . . . .	31
1.4.3. Faut-il absolument migrer vers une architecture big data ? . . . . .	33
1.5. Enjeux et opportunités pour le monde de l'assurance . . . . .	34
1.6. Conclusion . . . . .	36
1.7. Bibliographie. . . . .	37

**Chapitre 2. Des méthodes classiques d'analyse des données au big data** . . . . . 41

Gilbert SAPORTA

- 2.1. De l'analyse des données au *data mining* : explorer et prédire. . . . . 41
- 2.2. Des approches obsolètes. . . . . 43
- 2.3. Comprendre ou prédire ? . . . . . 44
- 2.4. Validation des modèles prédictifs . . . . . 45
  - 2.4.1. Eléments de théorie de l'apprentissage . . . . . 45
  - 2.4.2. La validation croisée . . . . . 48
- 2.5. Combinaison de modèles . . . . . 48
- 2.6. Le cas de la grande dimension . . . . . 50
  - 2.6.1. Les régressions régularisées . . . . . 50
    - 2.6.1.1. La régression sur composantes principales . . . . . 50
    - 2.6.1.2. La régression PLS. . . . . 51
    - 2.6.1.3. La régression *ridge* . . . . . 52
  - 2.6.2. Les méthodes *sparse* . . . . . 52
    - 2.6.2.1. Le Lasso . . . . . 52
    - 2.6.2.2. ACP et ACM *sparse* . . . . . 53
- 2.7. La fin de la science ? . . . . . 53
- 2.8. Bibliographie . . . . . 54

**Chapitre 3. Méthodes d'apprentissage statistique** . . . . . 57

Franck VERMET

- 3.1. Introduction . . . . . 57
  - 3.1.1. Apprentissage supervisé . . . . . 58
  - 3.1.2. Apprentissage non supervisé . . . . . 59
- 3.2. Les arbres de décision . . . . . 60
- 3.3. Les réseaux de neurones. . . . . 63
  - 3.3.1. Du neurone réel au neurone formel . . . . . 64
  - 3.3.2. Le Perceptron simple comme séparateur linéaire . . . . . 65
  - 3.3.3. Le Perceptron multicouche comme outil d'approximation de fonctions . . . . . 67
  - 3.3.4. L'algorithme de rétropropagation du gradient . . . . . 69
- 3.4. Les machines à vecteurs supports (SVM). . . . . 75
  - 3.4.1. Séparateur linéaire . . . . . 75
    - 3.4.1.1. Hyperplan séparateur de marge maximale. . . . . 75

3.4.1.2. Le cas non linéairement séparable . . . . .	78
3.4.2. Séparateur non linéaire . . . . .	78
3.5. Les méthodes d'agrégation de modèles . . . . .	79
3.5.1. <i>Bagging</i> . . . . .	80
3.5.2. Forêts aléatoires . . . . .	81
3.5.3. <i>Boosting</i> . . . . .	82
3.5.4. <i>Stacking</i> . . . . .	85
3.6. L'algorithme de classification non supervisée de Kohonen . . . . .	86
3.6.1. Notations et définition du modèle . . . . .	87
3.6.2. L'algorithme de Kohonen . . . . .	89
3.6.3. Applications . . . . .	90
3.7. Bibliographie . . . . .	91

## **Chapitre 4. Vision actuelle et prospective du marché . . . . . 95**

Florence PICARD

4.1. Le marché de l'assurance : structuré, réglementé et inscrit dans le long terme . . . . .	95
4.1.1. Une profession très réglementée et contrôlée . . . . .	96
4.1.2. Des activités très variées, inscrites dans le long terme . . . . .	97
4.1.3. Un marché lié à l'activité économique . . . . .	99
4.1.4. Des produits qui sont des contrats : un métier basé sur le droit . . . . .	99
4.1.5. Un modèle économique qui repose sur les données et l'expertise actuarielle . . . . .	100
4.2. Le contexte big data : nouveaux usages, nouveaux comportements et nouveaux modèles économiques . . . . .	101
4.2.1. Impact du big data sur les sociétés d'assurance . . . . .	101
4.2.2. Big data et numérique : un profond changement sociétal . . . . .	102
4.2.3. Confiance des clients dans les algorithmes et la technologie . . . . .	104
4.2.4. Une certaine insouciance quant aux conséquences possibles des traces numériques . . . . .	105
4.2.5. De nouveaux modèles économiques . . . . .	106
4.3. Des opportunités : nouvelles méthodes, nouvelles offres, nouveaux risques assurables, nouveaux outils de gestion . . . . .	107
4.3.1. De nouvelles méthodes de traitement des données . . . . .	107
4.3.2. Un marketing personnalisé et des tarifs affinés . . . . .	109
4.3.3. De nouvelles offres basées sur de nouveaux critères . . . . .	111
4.3.4. De nouveaux risques à assurer . . . . .	112
4.3.5. De nouvelles méthodes pour mieux servir les clients et mieux gérer . . . . .	113

4.4. Des risques de fragilisation de l'activité : concurrence de nouveaux acteurs, « ubérisation », contraction de volume du marché . . .	114
4.4.1. Le risque de démutualisation . . . . .	114
4.4.2. Le risque d'« ubérisation » . . . . .	115
4.4.3. Le risque d'un « Google » omniscient, en position dominante grâce aux data . . . . .	115
4.4.4. Le risque de concurrence avec des sociétés nouvelles créées pour le digital . . . . .	116
4.4.5. Le risque d'une réduction du champ des assurances de biens . . . .	117
4.4.6. Le risque de nonaccès aux données ou d'interdiction d'utilisation . . . . .	118
4.4.7. Le risque de cyberattaques et le risque de non-conformité. . . . .	119
4.4.8. Des risques de rigidités internes et des efforts de formation à accomplir. . . . .	120
4.5. Les enjeux de l'éthique et de la confiance . . . . .	120
4.5.1. Charte éthique et labellisation : preuves de loyauté . . . . .	121
4.5.2. Tarif, éthique et confiance . . . . .	122
4.6. Mobilisation des assureurs face au big data. . . . .	123
4.6.1. Une première phase « nouveaux convertis ». . . . .	124
4.6.2. Une phase d'appropriation et d'expérimentation dans différents domaines. . . . .	125
4.6.3. Des évolutions d'organisation et de management et de gros efforts de formation à effectuer . . . . .	128
4.6.4. Une nouvelle forme d'assurance : assurance « connectée » . . . . .	129
4.6.5. Insurtech et économie collaborative font pression pour l'innovation .	131
4.7. Pistes de stratégie pour l'avenir . . . . .	132
4.7.1. Paradoxes et difficultés d'anticipation . . . . .	132
4.7.2. Plusieurs choix envisageables . . . . .	133
4.7.3. Des évolutions incontournables . . . . .	137
4.8. Bibliographie . . . . .	138

**Chapitre 5. Utilisation du *big data* en assurance . . . . . 139**

Emmanuel BERTHELÉ

5.1. L'assurance, un secteur particulièrement adapté au développement du big data . . . . .	139
5.1.1. Un secteur qui s'est développé grâce à l'usage des données. . . . .	139
5.1.1.1. Secteur informatisé de longue date . . . . .	139
5.1.1.2. Secteur dont la donnée est la matière première . . . . .	141
5.1.1.3. Secteur dont le cadre réglementaire est incitatif en la matière . . . . .	142

---

5.1.2. Lien entre données et matière assurable . . . . .	143
5.1.2.1. Assurabilité . . . . .	143
5.1.2.2. Notion de mutualisation . . . . .	144
5.1.2.3. Le fantasme du tarif individualisé. . . . .	144
5.1.2.4. Adaptation de l'assurance aux usages grâce aux données disponibles . . . . .	145
5.1.3. Multiplication des sources de données présentant un intérêt potentiel . . . . .	146
5.1.3.1. Données directement disponibles pour l'assureur . . . . .	146
5.1.3.2. Données des objets connectés . . . . .	147
5.1.3.3. Données des réseaux sociaux . . . . .	147
5.1.3.4. Données externes mises à disposition . . . . .	148
5.2. Exemples d'application dans différentes activités d'assurance. . . . .	148
5.2.1. Utilisation à des fins de tarification et d'orientation de l'offre produits . . . . .	149
5.2.2. Assurance automobile et <i>télématique</i> . . . . .	151
5.2.3. Assurance indicielle d'événements météo-sensibles . . . . .	152
5.2.4. Orientation de l'épargne en assurance vie dans un contexte de taux bas . . . . .	154
5.2.5. Lutte contre la fraude . . . . .	155
5.2.6. Gestion d'actifs . . . . .	157
5.2.7. Réassurance . . . . .	157
5.3. Nouveaux métiers et évolutions des organisations induites pour les sociétés d'assurance. . . . .	158
5.3.1. Nouveaux métiers liés à la gestion, au traitement et à la valorisation des données . . . . .	158
5.3.2. Développement de partenariats entre assureurs et entreprises tierces. . . . .	159
5.4. Contraintes de développement . . . . .	160
5.4.1. Contraintes spécifiques au secteur de l'assurance . . . . .	160
5.4.1.1. Contraintes liées à la réglementation du secteur de l'assurance . . . . .	160
5.4.1.2. Contraintes liées à la CNIL . . . . .	160
5.4.1.3. Contraintes liées aux priorités des acteurs . . . . .	161
5.4.2. Contraintes non spécifiques au secteur de l'assurance . . . . .	162
5.4.2.1. Freins réglementaires . . . . .	162
5.4.2.2. Freins culturels . . . . .	164

5.4.3. Contraintes, en fonction des finalités, quant aux types d'algorithmes utilisables . . . . .	165
5.4.4. Rareté des profils et différences principales avec les actuaires . . .	166
5.5. Bibliographie . . . . .	168
<b>Liste des auteurs.</b> . . . . .	<b>169</b>
<b>Index</b> . . . . .	<b>171</b>